Replications of replications suggest that prior failures to replicate were not due to failure to replicate well

Center for Open Science, Charlottesville, VA

Critics said that a well-known psychology replication project failed to replicate findings because the replications had problems. Replications of the replications suggest otherwise.

The Reproducibility Project: Psychology (RP:P; Open Science Collaboration, 2015, *Science*) produced a surprising, even alarming, result. Attempts to replicate 100 published findings in psychology succeeded less than 40% of the time. This prompted global interest and debate about the credibility of psychological research. One critique (Gilbert et al., 2016, *Science*) asserted that some failures to replicate were a consequence of inadequate sample size (low power) and the replicators' failure to adhere to experts' insight for designing the replication studies. A team of 171 researchers tested whether these arguments had merit by conducting new replications of the replications. Today, the project is published as 11 articles comprising the entire Fall issue of the *Association for Psychological Science* journal *Advances in Methods and Practices in Psychological Science*. The findings cast doubt on Gilbert and colleagues' conclusions. A dramatic increase in sample size and expert peer review of 10 replication designs before conducting the studies did not increase replicability of the original findings. If the original findings are replicable, then the conditions necessary to observe them are not yet understood.

The team examined 10 of the 11 findings from the RP:P that had been labeled by replication teams as "not endorsed" by original authors.[1] These were studies in which the original authors had expressed reservations about the replication methodology that the original replication team did not completely address. The Many Labs 5 replication teams then revised the protocol to improve adherence to expert advice, and submitted the protocol to formal peer review before conducting the study at the journal. Replication teams addressed reviewer feedback until the revised protocol was accepted by the editor and then preregistered. Then, replication teams administered the revised replication protocol and the RP:P replication protocol to samples in multiple laboratories. This way, there would be a direct comparison of whether the expert feedback improved replicability of the original findings. The protocols were administered in 3 to 9 laboratories (median 6.5) to a total sample size of 276 to 3512 (median 1279.5), more than 16x larger than the original studies that generated the novel findings (median 76).

---

[1] The team could not recruit a sufficient number of labs to conduct replications of the 11th to include it in the investigation.

Two formal analysis strategies for testing whether the Revised protocol improved replicability compared to the RP:P protocol failed to find robust evidence of improvement. Descriptively, the median effect size for the Revised replication protocol (r = .05) was similar to the RP:P replication protocol (r = .04) and the original RP:P replications (r = .11). And, all of them were smaller than the original studies (r = .37). Charlie Ebersole, lead author of the project and Postdoctoral Associate at the University of Virginia, said "We tested whether revising the replication protocols based on expert reviews could improve replicability of the findings, and we found that it had no meaningful impact on these findings. Overall, the effects generated by the original replications were very similar to those generated by our revised protocols. Looking at all of these replications, our evidence suggests that the original studies may have exaggerated the existence or size of the findings." Added co-author Christopher Chartier, Associate Professor of Psychology at Ashland University, "If the original findings are credible, the conditions necessary for obtaining them are not yet known."

Hans IJzerman, co-author and Associate Professor at Université Grenoble Alpes, noted that "These results do not suggest that expertise is irrelevant. It could be that this particular selection of studies--ones that had already failed to replicate--were unlikely to improve no matter what expert feedback was provided. It will be interesting to conduct follow-up research on findings that are known to be replicable but have complex methodologies to help assess the role of expertise in achieving replicable results." Hugh Rabagliati, co-author and Reader in Psychology at Edinburgh University added "There were hints that some of the findings may be replicable, and perhaps even slightly more so with the revised protocols for one or two of them. However, overall, the cumulative evidence was 78% smaller than the original studies alone on average. And, because we had very large samples, our findings had much more precision than the original studies."

The findings are evidence against the hypothesis that the earlier failures to replicate these 10 studies were due to deficiencies in power and adherence to expert feedback. Meta-analyses combining the original finding and all replication studies indicated just 4 having statistically significant results (p < .05) and 3 of those weakly so. Future research may identify still conditions that improve replicability of these findings. "For now, the cumulative evidence suggests that the effects are weaker than original suggests or not yet established as a reliable finding," concluded Erica Baranski, co-author and Postdoctoral Researcher at the University of Houston.

The original findings were published in 2008. Since recognition of replicability challenges, the field of psychology has undergone substantial changes in its research practices to improve rigor and transparency, with the presumption that it will likewise improve replicability. For example,
- Many psychology journals have updated their policies to improve transparency and promote reproducibility, such as adoption of Registered Reports (http://cos.io/rr/). For example, the most progressive journals as rated by TOP Factor (http://cos.io/top/; http://topfactor.org/) and index of transparency and openness policies are heavily represented by psychology journals.

- A metascience research community in psychology emerged as a grassroots effort. This movement is self-scrutinizing the field's practices and findings and testing innovations to improve research practices. An example is the Meta-Research Center at Tilburg School of Social and Behavioral Sciences (https://metaresearch.nl/) that arose in the aftermath of the Dederick Stapel fraud case at Tilburg.
- Recent surveys of researcher behaviors suggest a dramatic increase in psychologists behaviors to improve rigor and transparency, particularly in increasing rates of preregistrations and sharing data, materials, and code (https://osf.io/preprints/metaarxiv/5rksu/). The Open Science Framework (http://osf.io), for example, now has more than 260,000 registered users that have registered more than 50,000 studies and shared more than 8 million files. About 4 million researchers will have accessed and downloaded that shared content, just in 2020. OSF is used by researchers from all disciplines, but had its origins and initial adoption by psychologists seeking to change the research culture toward openness and reproducibility.

"If psychology's reform continues to improve the transparency and rigor of research, I expect that future replication efforts will demonstrate the tangible impact of those improvements on research credibility," concluded Brian Nosek, senior author and Executive Director of the Center for Open Science.

Reference Materials

Summary meta-analysis paper for Many Labs 5
Ebersole, C. R., Mathur, M. B., Baranski, E., Bart-Plange, D., Buttrick, N. R., Chartier, C. R., Corker, K. S., Corley, M., Hartshorne, J. K., IJzerman, H., Lazarevic, L. B., Rabagliati, H., Ropovik, I., Aczel, B., Aeschbach, L. F., Andrighetto, L., Arnal, J. D., Arrow, H., Babincak, P., Bakos, B. E., Baník, G., Baskin, E., Belopavlović, R., Bernstein, M. H., Białek, M., Bloxsom, N. G., Bodroža, B., Bonfiglio, D. B. V., Boucher, L., Brühlmann, F., Brumbaugh, C., Casini, E., Chen, Y., Chiorri, C., Chopik, W. J., Christ, O., Ciunci, A. M., Claypool, H. M., Coary, S., Čolić, M. V., Collins, W. M., Curran, P. G., Day, C. R., Dering, B., Dreber, A., Edlund, J. E., Falcão, F., Fedor, A., Feinberg, L., Ferguson, I. R., Ford, M., Frank, M. C., Fryberger, E., Garinther, A., Gawryluk, K., Gerken, K., Giacomantonio, M., Giessner, S. R., Grahe, J. E., Guadagno, R. E., Hałasa, E., Hancock, P. J. B., Hilliard, R. A., Hüffmeier, J., Hughes, S., Idzikowska, K., Inzlicht, M., Jern, A., Jiménez-Leal, W., Johannesson, M., Joy-Gaba, J. A., Kauff, M., Kellier, D. J., Kessinger, G., Kidwell, M. C., Kimbrough, A. M., King, J. P. J., Kolb, V. S., Kołodziej, S., Kovacs, M., Krasuska, K., Kraus, S., Krueger, L. E., Kuchno, K., Lage, C. A., Langford, E. V., Levitan, C. A., de Lima, T. J. S., Lin, H., Lins, S., Loy, J. E., Manfredi, D., Markiewicz, Ł., Menon, M., Mercier, B., Metzger, M., Meyet, V., Millen, A. E., Miller, J. K., Moore, D. A., Muda, R., Nave, G., Nichols, A. L., Novak, S. A., Nunnally, C., Orlić, A., Palinkas, A., Panno, A., Parks, K. P., Pedović, I., Pękala, E., Penner, M. R., Pessers, S., Petrović, B., Pfeiffer, T., Pieńkosz, D., Preti, E., Purić, D., Ramos, T., Ravid, J., Razza, T. S., Rentzsch, K., Richetin, J., Rife, S. C., Rosa, A. D., Rudy, K. H., Salamon, J., Saunders, B., Sawicki, P., Schmidt, K., Schuepfer, K., Schultze, T., Schulz-Hardt, S., Schütz, A., Shabazian, A., Shubella, R. L., Siegel, A., Silva, R., Sioma, B., Skorb, L., de Souza, L. E. C., Steegen, S., Stein, LAR, Sternglanz, R. W., Stojilović,

D., Storage, D., Sullivan, G. B., Szaszi, B., Szecsi, P., Szoke, O., Szuts, A., Thomae, M., Tidwell, N. D., Tocco, C., Torka, A., Tuerlinckx, F., Vanpaemel, W., Vaughn, L. A., Vianello, M., Viganola, D., Vlachou, M., Walker, R. J., Weissgerber, S. C., Wichman, A. L., Wiggins, B. J., Wolf, D., Wood, M. J., Zealley, D., Žeželj, I., Zrubka, M., & Nosek, B. A. (2020). Many Labs 5: Testing pre-data collection peer review as an intervention to increase replicability. *Advances in Methods and Practices in Psychological Science, 3*, XXX-XXX.

Ten Many Labs 5 papers reporting detail of each investigation

1. Baranski, E., Baskin, E., Coary, S., Ebersole, C. R., Krueger, L. E., Lazarevic´, L. B., . . . Žeželj, I. (2020). Many Labs 5: Registered Replication of Shnabel and Nadler (2008), Study 4. *Advances in Methods and Practices in Psychological Science*, 3, XXX–XXX.
2. Buttrick, N. R., Aczel, B., Aeschbach, L. F., Bakos, B. E., Brühlmann, F., Claypool, H. M., . . . Wood, M. J. (2020). Many Labs 5: Registered Replication of Vohs and Schooler (2008), Experiment 1. *Advances in Methods and Practices in Psychological Science, 3*, XXX–XXX.
3. Chartier, C. R., Arnal, J. D., Arrow, H., Bloxsom, N. G., Bonfiglio, D. B. V., Brumbaugh, C. C., . . . Tocco, C. (2020). Many Labs 5: Registered Replication of Albarracín et al. (2008), Experiment 5. *Advances in Methods and Practices in Psychological Science, 3*, XXX–XXX.
4. Corker, K. S., Arnal, J. D., Bonfiglio, D. B. V., Curran, P. G., Chartier, C. R., Chopik, W. J., . . . Wiggins, B. J. (2020). Many Labs 5: Registered Replication of Albarracín et al. (2008), Experiment 7. *Advances in Methods and Practices in Psychological Science, 3*, XXX–XXX.
5. Ebersole, C. R., Andrighetto, L., Casini, E., Chiorri, C., Dalla Rosa, A., Domaneschi, F., . . . Vianello, M. (2020). Many Labs 5: Registered Replication of Payne, Burkley, and Stokes (2008), Study 4. *Advances in Methods and Practices in Psychological Science*, 3, XXX–XXX.
6. IJzerman, H., Ropovik, I., Ebersole, C. R., Tidwell, N. D., Markiewicz, Ł., Souza de Lima, T. J., . . . Day, C. R. (2020). Many Labs 5: Registered Replication of Förster, Liberman, and Kuschel's (2008) Study 1. *Advances in Methods and Practices in Psychological Science, 3,* XXX–XXX.
7. Lazarevic´, L.B., Puric´, D., Žeželj, I., Belopavlovic´, R., Bodroža, B., Cˇolic´, M. V., . . . Stojilovic´, D. (2020). Many Labs 5: Registered Replication of LoBue and DeLoache (2008). *Advances in Methods and Practices in Psychological Science, 3,* XXX–XXX.
8. Mathur, M. B., Bart-Plange, D.-J., Aczel, B., Bernstein, M. H., Ciunci, A. M., Ebersole, C. R., . . . Frank, M. C. (2020). Many Labs 5: Replication of the tempting-fate effects in Risen and Gilovich (2008). *Advances in Methods and Practices in Psychological Science, 3,* XXX–XXX.
9. Rabagliati, H., Corley, M., Dering, B., Hancock, P. J. B., King, J. P. J., Levitan, C. A., . . . Millen, A. E. (2020). Many Labs 5: Registered Replication of Crosby, Monin, and Richardson (2008). *Advances in Methods and Practices in Psychological Science, 3*, XXX–XXX.

10. Skorb, L., Aczel, B., Bakos, B. E., Feinberg, L., Hałasa, E., Kauff, M., . . . Hartshorne, J. K. (2020). Many Labs 5: Replication of van Dijk, van Kleef, Steinel, and van Beest (2008). *Advances in Methods and Practices in Psychological Science, 3,* XXX–XXX.

The preregistrations, data, materials, and code for all of the replications are publicly available on the Open Science Framework at https://osf.io/7a6rd/.

Reproducibility Project: Psychology (2015). https://science.sciencemag.org/content/349/6251/aac4716.long

D. Gilbert et al. critique (2016). https://science.sciencemag.org/content/351/6277/1037.2.abstract

Replies to D. Gilbert critique (2016) by
- Reproducibility Project authors. https://science.sciencemag.org/content/351/6277/1037.3
- B. Nosek & E. Gilbert. https://psyarxiv.com/nt4d3/

Contacts for inquiries about Many Labs 5

Charlie Ebersole	cebersole@virginia.edu
Brian Nosek	nosek@virginia.edu
Maya Mathur	mmathur@stanford.edu
Katie Corker	corkerka@gvsu.edu
Hans IJzerman	h.ijzerman@gmail.com
Hugh Rabagliati	hugh.rabagliati@ed.ac.uk
Nick Buttrick	nrb8pv@virginia.edu
Lily Lazarević	ljiljana.lazarevic@f.bg.ac.rs
Christopher Chartier	cchartie@ashland.edu
Erica Baranski	ericanbaranski@gmail.com

About Center for Open Science

The Center for Open Science (COS) is a non-profit technology and culture change organization founded in 2013 with a mission to increase openness, integrity, and reproducibility of scientific research. COS pursues this mission by building communities around open science practices, supporting metascience research, and developing and maintaining free, open source software tools. The OSF is a web application that provides a solution for the challenges facing researchers who want to pursue open science practices, including: a streamlined ability to manage their work; collaborate with others; discover and be discovered; preregister their studies; and make their code, materials, and data openly accessible. Learn more at cos.io and osf.io.

Contact for the Center for Open Science
Inquiries: Claire Riss claire@cos.io
Web: cos.io
Twitter: @osframework